AD-A091 516 SOUTHERN METHODIST UNIV DALLAS TEX DEPT OF STATISTICS F/G 12/1
THE DETERMINATION OF SEASONAL ARMA MODELS.(U)
OCT 80 M MORTON N00014-75-C-0439
UNCLASSIFIED TR-141 NL

1 of 1
AD A
091516

END
DATE
FILMED
12-80
DTIC

(4)

LEVEL

AD A091516

4

(6) THE DETERMINATION OF SEASONAL ARMA MODELS

by

(10) Michael Morton

(9) Technical Rep~~ort~~, No. 141
Dep~~artment of Statistics~~ ONR Contract

(12) 35

(11) October 1980

(14) TR-141

Research sponsored by ~~the Office of Naval~~ Research
(15) ~~Contract~~ N00014-75-C-0439
Project ~~NR 042-280~~

Reproduction in whole or in part is permitted
for any purpose of the United States Government

This document has been approved for public
release and sale; its distribution is unlimited

DIVISION OF MATHEMATICAL SCIENCES
Department of Statistic
Southern Methodist University
Dallas, Texas 75275

404198

dl

THE DETERMINATION OF SEASONAL ARMA MODELS

by

Michael Morton

SECTION I

INTRODUCTION

Gray, Kelley, and McIntire [1978] have described a method for determining the order of an ARMA process and for identifying roots of the characteristic equation on or near the unit circle. In this paper, we will demonstrate how that approach can be utilized in modeling seasonal data.

Many practitioners, at present, perfunctorily employ the operator $1-B^S$ on any data set believed to have a period of length S. Use of that operator, however, tacitly assumes not only a frequency of 1/S to be present in the data, but of all the harmonics of 1/S (i.e., it assumes the frequencies 0, 1/S, 2/S,..., $\frac{[S/2]}{S}$ where [.] is the greatest integer function). We shall demonstrate a technique which will aid in determining when such an operator is called for and when other seasonal models are called for.

In the next section, we define our terms and give the theorems which are necessary for describing the method which we employ. In the following section, we illustrate the procedure using two example series: the International Airline series given in Box and Jenkins [1976] and the so-called Radio series given by Siddiqui [1962].

Accession For
Justification
By
Distribution/
Availability Codes

SECTION II

DEFINITIONS AND THEOREMS

Definition 1

By an ARMA (p,q) process, we mean a stochastic process $\{X_t\}$ which satisfies

$$\phi(B)\, X_t = \theta(B) a_t,\ t = 0, \pm 1, \pm 2, \ldots$$

where

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p,$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q \text{ with}$$

$$\phi_p, \theta_q \neq 0 \text{ and } B^k X_t = X_{t-k}.$$

The algebraic equation $\phi(r) = 0$ is called the characteristic equation of the corresponding ARMA process. We assume that $\phi(r)$ has all of its roots on or outside the unit circle, and that $\phi(r)$ and $\theta(r)$ are relatively prime. $\{a_t\}$ is assumed to be a white noise process.

It is well known that $\{X_t\}$ is a stationary process if, and only if all of the roots of its characteristic equation are strictly outside the unit circle. By a non-stationary ARMA process, we will mean an ARMA process with one or more of the roots of $\phi(r)$ lying on the unit circle. That is, we exclude the case of roots inside the unit circle. The term seasonal model will be used to designate the following class of non-stationary ARMA processes.

Definition 2

A factor $\Psi(B)$ will be said to be seasonal if

$$\Psi(B) = 1 + B, \quad \text{or}$$

$$\Psi(B) = 1 + \Psi_1 B + B^2, \ |\Psi_1| < 2.$$

An ARMA (p,q) process will then be referred to as seasonal if it consists of one or more seasonal factors.

Motivation for the above definition is seen most easily in the frequency domain. A seasonal factor is any (irreducible) non-stationary factor with associated frequency greater than 0.

Definition 3

The autocorrelation of a stationary ARMA process is given by

$$\rho(k) = E(X_t X_{t+k}) \;/\; E(X_t^2).$$

Strictly speaking the autocorrelation of a non-stationary ARMA process does not exist. However, if one regards a non-stationary ARMA process as a limiting case of a sequence of stationary processes, the following definition will appear natural.

Definition 4

Let $\rho_k\,(\lambda_1,\ldots,\lambda_p,\underline{\theta})$ denote the autocorrelation at lag k of a stationary ARMA process with $\lambda_1,\ldots,\lambda_p$, the roots of $\phi(r)$ and $\underline{\theta}' = (\theta_1,\ldots,\theta_q)$ the moving average parameters. Now suppose that $\{X_t\}$ is a non-stationary ARMA (p,q) process with roots of $\phi(r)$ $\lambda_1,\ldots,\lambda_p$ of which $\lambda_1,\ldots,\lambda_m$ are on the unit circle. We then extend the definition of $\rho(k)$ by letting

$$\rho(k) = \lim_{\alpha\to 1^+} \rho_k(\alpha\lambda_1,\alpha\lambda_2,\ldots,\alpha\lambda_m,\lambda_{m+1},\ldots,\lambda_p,\underline{\theta})$$

We will call $\rho(k)$ the autocorrelation of the non-stationary process $\{X_t\}$.

Definition 5

If $X_1,\ldots,X_n$ are consecutive random variables from the ARMA process $\{X_t\}$, we take as our estimator of $\rho(k)$

$$\hat{\rho}(k) = \sum_{t=1}^{n-|k|} (X_t - \bar{X})(X_{t+|k|} - \bar{X}) \Big/ \sum_{1}^{n} (X_t - \bar{X})^2$$

Fundamental to the discussion to follow is the so-called S-array. For completeness, we give a formal definition of the S-array; however, its importance for our purposes is contained in the two theorems which follow. (Our definition differs slightly from that normally given. It represents a simple shift in index in order to give the format of the shifted S-array suggested by Woodward and Gray [1979] in simpler notation).

Definition 6

Given a doubly infinite sequence $\{f_m\}$ let

$$S_n(f_m) = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ f_{m-n+1} & f_{m-n+2} & \cdots & f_{m+1} \\ \vdots & \vdots & & \vdots \\ f_m & f_{m+1} & \cdots & f_{m+n} \end{vmatrix} \Bigg/ \begin{vmatrix} f_{m-n+1} & \cdots & f_m \\ \vdots & & \vdots \\ f_m & \cdots & f_{m+n-1} \end{vmatrix}$$

The S-array is then the numbers $S_n(f_m)$ displayed as in Table 1.

TABLE 1

m/n	1	...	k
$-\ell$	$S_1(-\ell)$	...	$S_k(-\ell)$
$-\ell+1$	$S_1(-\ell+1)$	...	$S_k(-\ell+1)$
$\vdots$	$\vdots$		$\vdots$
-1	$S_1(-1)$	...	$S_k(-1)$
0	$S_1(0)$	...	$S_k(0)$
1	$S_1(1)$	...	$S_k(1)$
2	$S_1(2)$	...	$S_k(2)$
$\vdots$	$\vdots$		$\vdots$
j	$S_1(j)$	...	$S_k(j)$

$S_n(f_m)$

Theorem 1

If $\{X_t\}$ is a stationary ARMA (p,q) process and if $f_m = \rho(m)$ or $f_m = (-1)^m \rho(m)$, then $S_n(f_m) = C_1$ for all $m \geq m_1$ and $S_n(f_m) = C_2$ for all $m \leq m_2$ if, and only if $n = p$, $m_1 = q$, and $m_2 = -q-1$, where C_1 and C_2 are constants.

PROOF SEE GKM [1978]

We thus note that, given the true autocorrelation function of a stationary ARMA process, the S-array provides an unequivocal identification of p and q. Given an estimate $\hat{\rho}(k)$, we then look for a similar pattern in the S-array to provide information as to the order of the process $\{X_t\}$.

Theorem 2

If $\{X_t\}$ is an ARMA (p,q) process and if $f_m = \rho(m)$ or if $f_m = (-1)^m \rho(m)$ then $\{X_t\}$ is non-stationary if, and only if for some n and some C,

$$S_n(f_m) = C \text{ for all } m, \text{ where } C \text{ is a constant.}$$

In that case n is the number of roots of highest multiplicity among those roots of $\phi(r)$ located on the unit circle.

PROOF See Quinn [1980] and Theorem 1 GKM [1978]

Theorem 2 suggests that a stepwise procedure will be required for identifying the complete model whenever $\phi(r)$ has roots on the unit circle. First, the presence of non-stationarities are detected by noting a column of the S-array which is relatively constant. The series is then transformed by the indicated non-stationary factor and the residual series is then investigated.

Few distributional properties of the S-array are known. However, Gray, Kelley, and McIntire [1978] have shown through a variety of examples that the S-array is relatively robust to stochastic disturbance. We also give two asympotic results.

Theorem 3

Suppose that $X_1, X_2, \ldots X_T$ are consecutive random variables from an ARMA (p,q) process satisfying $\phi(B)X_t = \theta(B)a_t$ and $r_T(k)$ is the sample autocorrelation function at lag k.

(i) Suppose $S_n(\rho(k))$ is defined and $\underset{T\to\infty}{P - \lim}\, r_T(k) = \rho(k)$, then

$$\underset{T\to\infty}{P\text{-}\lim}\, S_n(r_T(k)) = S_n(\rho(k))$$

if the roots of $\phi(r)$ are strictly outside the unit circle.

(ii) X_t is non-stationary if, and only if for some n and C

$$\underset{T\to\infty}{P - \lim}\, S_n(r_T(k)) = C$$

for all k, where C is independent of k, and $S_n(\rho(k))$ is defined.

PROOF

(i) This easily follows since S_n is a continuous function and $\underset{T\to\infty}{P - \lim}\, r_T(k) = \rho(k)$.

(ii) The proof of part (ii) relies on two quite useful results which were established by Findley [1980] and which we state as lemmas.

Lemma 1

Suppose all quantities are as defined in Theorem 3 and that

$$\phi(r) = [\prod_1^m (1-\alpha_i\, r)]^d\, \psi(r)$$

where the α_i are distinct, $|\alpha_i| = 1$ and $\psi(r)$ has no roots on the unit circle of multiplicity greater than d - 1. Writing

$$\prod_1^m (1-\alpha_i r) = 1 + a_1 r + \ldots + a_m r_m$$

we then have

$$P - \lim_{T\to\infty} \left(r_T(k) + a_1 r(k-1) + \ldots + a_m r_T(k-m) \right) = 0$$

for all k = 0, ± 1,....

Lemma 2 If we let:

$$H_m(r_T(k)) = \begin{vmatrix} r_T(k-m+1) & \cdots & r_T(k) \\ \vdots & & \vdots \\ r_T(k) & \cdots & r_T(k+m-1) \end{vmatrix},$$

the denominator quantity in the S-function, we have, taking m as in Lemma 1,

$$\liminf_T |H_m(r_T(k))| > 0$$

for all k, almost surely.

We may now prove part (ii) of Theorem 3 using the notation introduced above.

($\Rightarrow$) . Assume X_t is non-stationary and $\phi(r)$ is as defined in lemma 1. Fix k and let

$$U_T(i) = r_T(k + i) + a_1 r_T(k+i-1) + \ldots + a_m r_T(k+i-m)$$

for $i = 1, \ldots, m$.

By lemma 2, $\underset{T\to\infty}{P - \lim}\, U_T(i) = 0$, $i = 1, \ldots m$.

Further by performing simple column operations in the numerator determinant we have, letting A_{ij} be the ijth cofactor of the matrix in the numerator of the S-function,

$$S_m(r_T(k)) = (1 + a_1 + \ldots + a_m)\,(-1)^m + U_T(1) \quad A_{2,m+1}/H_m(r_T(k))$$
$$+ \ldots + U_T(m) \quad \frac{A_{m+1,m+1}}{H_m(r_T(k))}$$

Now, since A_{ij} is bounded it follows from lemmas 1 and 2 that:

$$\underset{T\to\infty}{P - \lim}\, S_m(r_T(k)) = (-1)^m(1 + a_1 + \ldots a_m)$$

($\Leftarrow$) Suppose then that there is an n so that

$$\underset{T\to\infty}{P - \lim}\, S_n(r_T(k)) = C \text{ for all } k$$

and that $S_n(\rho(k))$ is defined for all k.

If X_t is stationary, then with the above conditions:

$$\underset{T\to\infty}{P- \lim}\, S_n(r_T(k)) = S_n(\rho(k))$$

which is not the same for all k. Hence X_t is non-stationary and the theorem is proved.

SECTION III

ANALYSIS

To illustrate the usefulness of the S-array as a model identification tool, we consider two real data examples. Our first example is the International Airline Series given by Box and Jenkins [1976]. This example we will briefly examine even though it is analyzed in much the same manner by Gray and Woodward [1980] and by Hart and Gray [1980]. Our purpose for including the Airline Series is to show the contrast between it and our second example series: the so-called Radio Series, given by Siddiqui [1962].

The International Airline Series consists of the natural logarithm of the number of passengers in International air travel. The data are monthly values from January 1949 to December 1960.

A plot of the data (see figure 1) shows that it appears to have a linear trend and a quite distinctive yearly periodicity about that trend. The S-array, evaluated using $f_m = (-1)^m\hat{\rho}(m)$, is shown in Table 2.

An ambiquity regarding the identification is noted, since both the 1st and 13th columns are relatively constant. The series will clearly not be well-modeled as a 1st order process; however the near constant 1st column indicates the presence of a near 1st order non-stationarity. It is usually best to remove that factor before attempting further identification. That factor is estimated as roughly 1-.95B, using the Yule-Walker estimate.

We next transformed by the operator given above. The S-array,

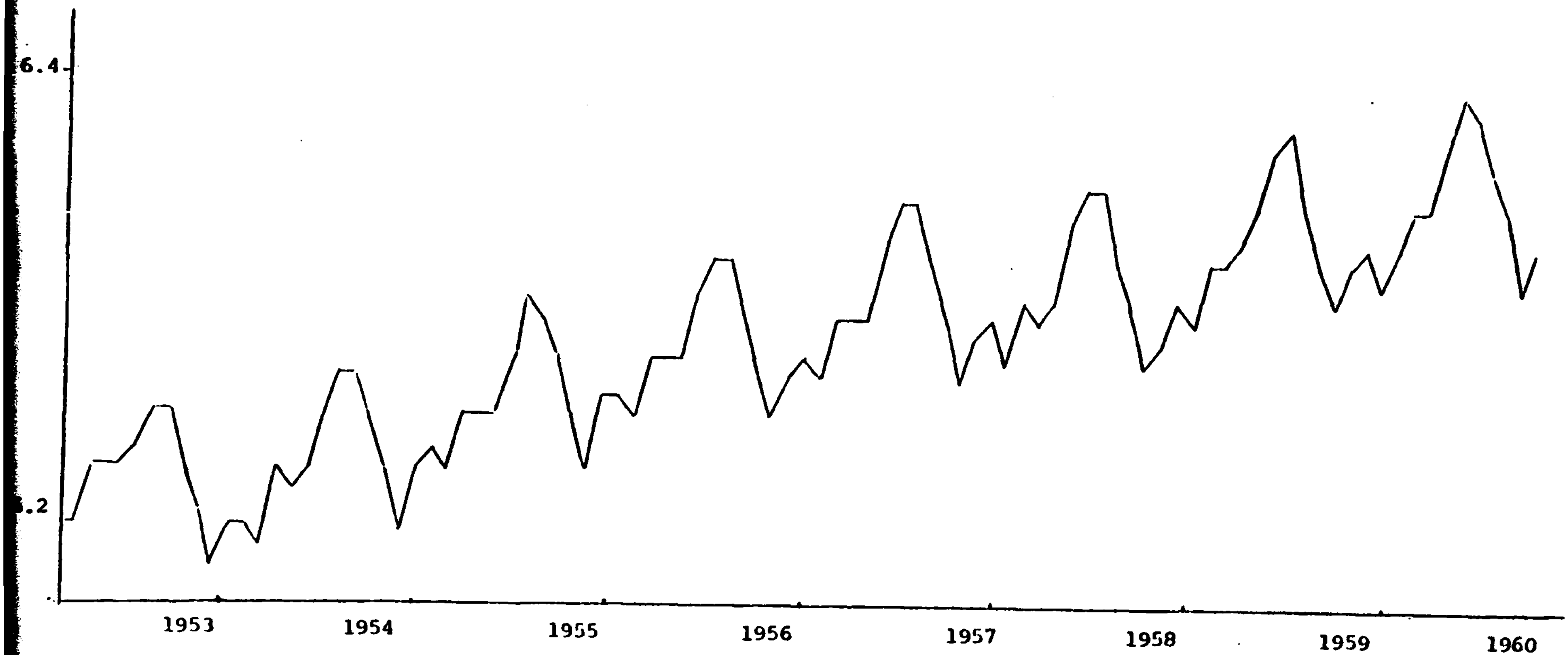

Figure 1

Airline Data

TABLE 2

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14
-12	-1.995	2.385	-2.633	2.499	-3.510	3.360	-3.750	3.086	-4.329	3.427	-5.308	4.938	10.060	-9.664
-11	-1.982	1.393	-3.729	21.070	-3.808	12.898	-8.780	11.177	-5.623	12.594	-5.712	******	10.293	******
-10	-1.986	4.663	-8.032	-6.513	-2.865	8.001	1.289	-42.149	-1.430	-2.257	-35.616	20.039	7.722	-3.679
- 9	-1.991	10.528	15.202	.266	-3.748	9.968	41.646	-48.707	1.229	15.062	-40.475	-49.521	6.993	7.580
- 8	-2.014	2.981	-3.106	3.828	-3.895	2.320	-1.057	-5.136	-29.858	29.780	-33.198	16.872	5.391	-6.092
- 7	-2.026	2.798	17.584	7.668	- .358	-.076	10.608	.206	-29.568	******	-58.766	-43.824	5.513	-35.207
- 6	-2.030	6.165	-5.477	12.773	.056	-9.041	11.472	30.163	-29.655	49.694	*******	35.123	4.746	- 7.228
- 5	-2.038	5.758	-91.139	6.362	10.397	-10.813	-1.151	50.555	-26.161	-80.379	-59.416	-62.153	5.050	-13.185
- 4	-2.052	2.663	- 1.721	-2.681	-18.935	47.713	-50.282	50.171	-.149	9.472	-20.459	15.143	4.482	1.853
- 3	-2.057	4.365	10.489	9.331	-29.006	38.563	37.763	******	-9.446	9.275	-23.463	-78.790	3.933	-.002
- 2	-2.061	-4.093	-30.833	55.979	-31.416	82.985	112.223	******	-19.338	25.536	-13.646	27.761	3.426	-.460
- 1	-2.049	18.571	-42.443	-94.592	-21.649	-54.007	-65.396	-22.480	-13.212	-39.497	-26.330	68.536	3.085	45.097
0	-1.954	2.183	-2.302	2.247	-2.507	2.396	-2.487	2.239	-2.697	2.524	-2.792	2.910	-1.498	1.549
1	-1.943	1.295	-3.255	13.463	-3.357	4.451	-8.754	7.285	-3.369	6.980	-1.792	34.767	-1.655	-.290
2	-1.946	3.688	-6.048	-2.624	-2.822	15.228	-4.762	59.831	-1.694	-1.562	-17.677	15.067	-1.870	.001
3	-1.950	8.564	5.575	1.914	-3.774	9.626	-41.300	23.564	-.029	10.689	-19.554	-35.587	-2.147	.688
4	-1.963	3.051	-3.143	4.308	-5.806	2.110	.052	-9.655	-10.521	10.501	-12.493	9.399	-2.351	3.561
5	-1.971	2.944	57.414	9.358	- .081	-.542	9.436	-9.810	-13.919	29.393	-18.384	-26.579	-2.197	7.291
6	-1.975	7.220	-5.346	18.708	.539	-1.999	13.541	-.239	-11.853	11.578	-74.994	18.459	-2.504	2.894
7	-1.986	6.195	-61.608	11.015	11.371	-13.444	12.538	11.770	-11.852	******	-43.211	-17.403	-2.441	25.674
8	-2.009	2.484	-8.480	- .195	-9.128	-14.982	-14.192	13.891	4.371	4.450	-16.839	10.081	-2.954	1.726
9	-2.014	3.515	11.312	8.480	-9.349	17.589	-2.014	168.806	-4.813	10.573	-18.321	-17.965	-3.319	-3.683
10	-2.019	-4.744	-18.937	30.603	-11.291	14.515	*******	146.726	-17.270	19.482	-13.964	11.931	-4.049	4.338
11	-2.005	11.863	-25.187	-51.183	-10.230	-76.770	-31.467	-17.057	-10.784	-16.824	-12.257	-78.161	-3.953	-60.721
12	-1.940	2.376	- 2.746	2.945	- 3.415	3.270	-3.668	3.096	-4.431	4.345	-5.863	5.893	-2.489	3.223

S-array for the untransformed Airline Data using $f_m=(-1)^m\hat{\rho}(m)$.

using $f_m = \hat{\rho}(m)$, for the transformed series is given in Table 3. There is now an unambiquous constancy in the 12th column. The Yule-Walker estimate of that operator is given in Table 4.

At this point a subjective decision must be made as to whether or not a seasonal model is desired. If a seasonal model is desired, it must be decided which factors to alter to the unit circle. From the factors given in Table 4, it is apparent that each of the frequencies associated with the operator $1-B^{12}$ is present. We also note, that, with the possible exception of the factor 1+.92B, all of the roots are near the unit circle.

Thus a reasonable approximation to the 2nd estimated operator is given by $1-B^{12}$. Likewise the operator 1-.95B, initially estimated, may be adjusted to the non-stationary operator 1-B. Other possibilities might be considered (see Gray and Woodward [1980]), but the work done so far indicates that the operator $(1-B)(1-B^{12})$ is not unreasonable.

Table 5 gives the S-array, using $f_m = (-1)^m\hat{\rho}(m)$, for the original series transformed by $(1-B)(1-B^{12})$. The residual series appears to be well-modeled by an AR(12). The Yule-Walker fit is given in Table 6. We thus arrive at the model

$$(1-B)(1-B^{12})\phi(B)X_t = a_t \qquad (1)$$

where $\{a_t\}$ is a white noise series with var a_t = .00136 and $\phi(B)$ is the stationary operator given in Table 6.

Taking $1-B^{12}$ as the only non-stationary component of the model, Gray and Woodward [1980] arrive at the model

$$(1-B^{12})\ \Phi(B)X_t = a_t \qquad (2)$$

TABLE 3

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14
-12	-.736	.875	-.901	1.245	-1.094	1.179	-1.288	1.983	-1.728	2.458	1.040	-.722	1.138	.719
-11	-1.409	.769	10.671	1.503	-1.925	-.015	5.551	2.194	-9.582	5.099	.536	2.840	1.243	-.907
-10	.021	-2.295	.164	1.427	3.324	-5.578	5.660	1.632	-3.532	18.038	-.943	-.359	1.265	-5.716
-9	2.282	-2.373	-1.523	1.402	5.170	-51.692	-4.023	2.010	4.333	7.476	2.848	-1.098	1.296	.181
-8	-.736	.928	-.764	-.048	-.339	-.508	-.827	4.945	-3.998	5.258	-2.244	-.436	1.096	-.528
-7	-1.700	1.803	1.520	.328	-.227	.021	-2.072	4.667	-16.649	8.097	6.119	-.769	1.154	2.519
-6	-1.893	-49.864	-1.478	-.327	-.007	2.372	-2.112	4.130	-6.109	23.042	-5.236	-.433	.766	-.479
-5	4.679	3.940	1.979	.218	-2.255	2.289	-1.759	3.893	16.270	11.184	8.379	-.568	.892	-4.046
-4	-.590	.473	-1.441	3.279	-3.432	3.567	-3.299	4.850	-3.146	5.218	-3.233	-.460	1.798	.197
-3	-.267	4.752	.108	4.970	-4.939	-.243	14.727	3.959	-10.186	6.827	10.218	-.682	1.101	-1.713
-2	-3.617	2.131	-5.230	4.970	*******	-14.989	16.442	4.861	-5.407	22.754	-5.672	-.465	.716	7.473
-1	3.503	6.291	-14.031	4.466	17.464	69.661	-9.137	4.764	66.947	10.331	7.042	-.539	6.031	2.769
0	-.778	.888	-.948	1.203	-1.125	1.144	-1.308	1.802	-1.755	2.114	-1.626	.405	-.379	.439
1	-1.382	.523	2.576	1.413	-1.413	-8.662	1.654	1.880	-15.814	4.011	4.884	.345	-1.396	.773
2	.364	-1.526	.074	1.411	220.279	-.146	-8.239	1.599	-3.058	14.708	-4.558	.489	-.847	1.455
3	1.440	-3.312	-1.442	1.396	3.115	8.103	-7.547	1.870	7.238	7.603	6.840	.322	-.519	.079
4	-.824	.969	-.782	-.139	-.307	-.615	-.598	3.919	-3.019	4.201	-2.589	.387	-.798	1.437
5	-2.120	2.339	1.405	.278	.009	-.180	-2.148	3.741	-10.311	6.044	5.320	.290	-1.120	-1.299
6	-2.428	-48.792	-1.534	-.438	.180	-.265	-1.829	3.340	-4.442	18.092	-3.797	.486	-.779	.597
7	2.788	3.354	2.466	.080	-2.620	1.810	-1.857	3.196	19.111	8.820	3.382	.260	-.871	-.734
8	-.695	.531	-1.360	2.990	-3.018	2.998	-2.662	4.522	-2.112	4.541	-1.595	.524	-.790	.112
9	-.020	3.480	.258	3.836	-2.693	36.863	4.532	3.092	-7.173	5.784	.835	.121	-.914	1.182
10	-3.448	3.357	-4.277	3.856	-14.894	-.066	-13.817	3.859	-4.651	15.136	-.437	.654	-.920	-2.328
11	2.792	4.585	-31.229	3.262	8.723	13.780	-13.626	3.587	13.309	4.592	-.778	3.022	-.865	.317
12	-.729	.896	-.900	1.199	-.984	1.159	-1.257	1.965	-1.363	2.616	1.329	.316	-.818	.606

S-array for the Airline Data after being transformed by (1-.95B)

TABLE 4

ESTIMATED WHITE NOISE VARIANCE .001922

ESTIMATED AR PARAMETERS .0767 -.0487 -.0175 -.0845 .0481 -.0327 -.0178 -.1139 .0646 -.0730 .0430 .7511

Root of Operator	Reciprocal of Root	Absolute Value of Root	Absolute Value of Reciprocal	Frequency	Period
(-.8915, .5025)	(-.8512, -.4798)	1.0234	.9771	.4183	2.3906
(-.8915, -.5025	(-8512, .4798)	1.0234	.9771	.4183	2.3906
(.8752, .5034)	(.8585, -.4938)	1.0097	.9904	.0831	12.0383
(.8752, -.5034)	(.8585, .4938)	1.0097	.9904	.0831	12.0383
(-1.0821, 0.0000)	(-.9242, 0.0000)	1.0821	.9242	.5000	2.000
(.5014, .8757)	(.4925, -.8600)	1.0091	.9910	.1672	5.9798
(.5014, -.8757)	(.4925, .8600)	1.0091	.9910	.1672	5.9798
(-.5093, .8764)	(-.4957, -.8529)	1.0137	.9865	.3338	2.9959
(-.5093, -.8764)	(-.4957, .8529)	1.0137	.9865	.3338	2.9959
(1.0401, 0.0000)	(.9615, 0.0000)	1.0137	.9615	0.0000	∞
(.0155, 1.0289)	(.0156, -.9716)	1.0291	.9717	.2474	4.0414
(.0165, -1.0289)	(.0156, .9716)	1.0291	.9717	.2474	4.0414

AR (12) fit to the Airline Series after being transformed by 1-.95B.

TABLE 5

S11	S12	S13	S14
-63.460	3.090	-1.861	.188
- 1.330	3.535	-3.286	-.270
- 2.188	3.514	-41.970	4.285
- 8.356	3.125	- 1.442	6.143
1.075	2.827	- 9.870	5.802
1.364	3.012	- .323	2.478
1.584	3.067	- 2.659	2.781
- 3.155	3.069	3.985	10.676
11.752	3.282	-11.955	24.985
-15.269	2.812	7.770	11.889
- .711	.952	- .848	.914
4.219	1.152	- 1.511	1.152
-5.292	1.096	- .601	-5.016
.031	1.064	6.888	-2.831
-1.613	1.057	.125	1.715
-1.536	1.036	-1.491	1.579
-1.430	1.064	1.107	1.825
1.781	1.292	-1.396	2.104
- .487	1.280	18.507	- .118
- .798	1.401	1.558	.095
27.383	1.028	.841	2.240

S-Array for the Airline Series after being transformed by $(1-B)(1-B^{12})$. $f_m=(-1)^m\hat{\rho}(m)$

TABLE 6

Estimated White Noise Variance .001364

Estimated AR Parameters -.3596 -.0528 -.1516 -.1092 .0473 .0883 -.0144 .0304 .1648 .0357 -.0805 -.3387

Root of Operator	Reciprocal of Root	Absolute Value of Root	Absolute Value of Reciprocal	Frequency	Period
(.7281, .7832)	(.6367, -.6849)	1.0694	.9351	.1308	7.6449
(.7281, -.7832)	(.6367, .6849)	1.0694	.9351	.1308	7.6449
(-.7928, .7573)	(-.6595, -.6300)	1.0964	.9121	.3786	2.6410
(-.7928, -.7573)	(-.6595, .6300)	1.0964	.9129	.3786	2.6410
(-.3849, 1.0680)	(-.2987, -.8287)	1.1352	.8809	.3051	3.2781
(-.3849, -1.0680)	(-.2987, -.8287)	1.1352	.8809	.3051	3.2781
(1.1180, .2838)	(.8403, -.2133)	1.1534	.8670	.0396	25.2783
(1.1180, -.2838)	(.8403, .2133)	1.1534	.8670	.0396	25.273
(.2395, 1.0281)	(.2149, -.9226)	1.0557	.9473	.2136	4.6823
(.2395, -1.0281)	(.2149, .9226)	1.0557	.9473	.2136	4.6823
(-1.0268, .2640)	(-.9135, -.2349)	1.0602	.9432	.4599	2.1742
(-1.0268, -.2640)	(-.9135, .2349)	1.0602	.9432	.4599	2.1742

AR(12) Yule-Walker fit to the Airline Series after being transformed by $(1-B)(1-B^{12})$.

where $\phi(B)$ is a 13th order stationary operator and $\{a_t\}$is a white noise process with var a_t = .001267.

Gray and Woodward [1979] discuss some of the considerations which are relevant in deciding between models (1) and (2). They argue for model (2) based on the models' respective forecast functions. They also give some comparison between the Box-Jenkins model, model (2) above, and a model given by Parzen [1979]. The reader is directed to the paper for further details.

The 2nd example we will consider is the so-called Radio Series given by Siddiqui [1962]. The data consists of the 240 monthly median f_0F_2 values observed in Washington, D.C. from May 1934 until April 1954.

A plot of the series and of the autocorrelation function (see Figures 2 and 3, respectively) each indicate the presence of a low frequency component and a quite distinctive yearly periodic oscillation. Noting the yearly period, many practitioners would apply the transformation $1-B^{12}$ to the data. Further analysis below, however, will show that operator to be unnecessary and in fact deleterious for this particular data set.

The S-array using $f_m=(-1)^m\rho(m)$ is given in Table 7. The 1st column is seen to be roughly constant, reflecting the low frequency component. Since the S-array is not too distinctive, in view of Theorem 2, it seems reasonable at this point to prefilter the data by 1-.9B. At this point, of course, we do not mean to imply that 1-.9B is a factor in the model but simply an appropriate high pass filter which allows clearer identification of the model.

The S-array for the transformed series, using $f_m= \rho(m)$ is given in Table 8. The S-array is relatively constant in the 13th column which indicates that the untransformed series should be well fit by an AR(14). The Yule-Walker fit is shown in Table 9. Note the

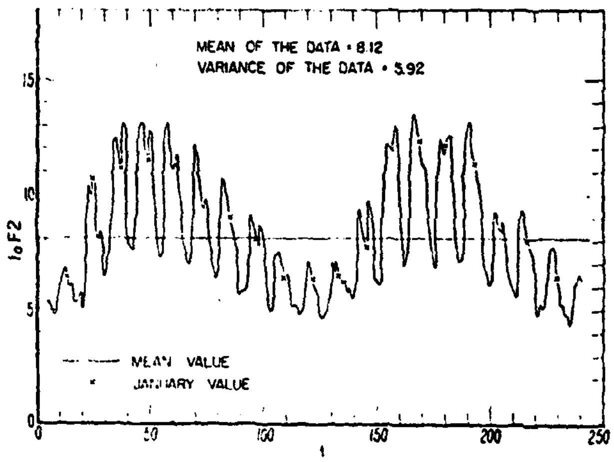

Figure 2. Plot of the first 236 points from the radio series

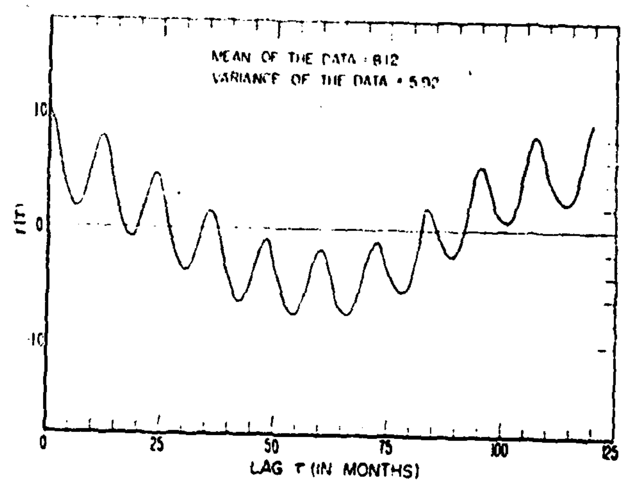

Figure 3. Plot of the estimated auto-correlation function of the radio series

TABLE 7

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14
-10	-1.737	2.302	-3.866	74.081	-14.509	-18.051	******	-74.487	-8.354	-44.527	-48.570	26.655	20.340	13.212
- 9	-.1707	1.745	-23.151	-8.474	- 6.252	- 2.816	3.897	- 7.413	-16.540	41.232	-92.514	-45.887	17.620	29.184
- 8	-1.706	*****	-10.293	3.539	- 3.283	- 3.445	-5.302	-31.319	- 9.052	-24.930	8.164	5.535	7.459	16.503
- 7	-1.884	4.948	- 6.699	2.984	40.747	- 4.398	28.745	-50.167	2.884	-17.764	-14.004	-3.143	1.574	14.234
- 6	-2.244	3.848	-10.363	-9.383	7.474	3.151	3.896	-11.984	23.702	-19.534	12.031	-1.132	-23.396	14.058
- 5	-2.491	1.882	- 3.984	-5.232	-35.180	- 5.357	7.022	- 7.688	-7.067	-23.785	-40.715	16.188	-16.551	13.716
- 4	-2.441	6.063	- 8.765	2.496	- 5.229	- 9.229	-8.740	-63.477	13.801	-27.001	- 1.061	25.407	-53.536	10.869
- 3	-2.392	10.303	8.112	8.140	-16.496	- 9.395	******	- 5.906	63.604	-23.034	-23.876	26.106	32.583	-.378
- 2	-2.276	6.409	- 9.396	-1.290	37.849	-11.645	22.928	-56.931	48.617	-12.231	- .340	8.833	9.539	8.021
- 1	-2.119	5.788	-58.894	-38.530	40.895	- 7.741	-15.650	- 8.499	-10.090	- 4.353	- 8.485	9.320	3.558	7.231
0	-1.893	2.814	- 2.955	2.745	- 2.572	1.931	- 2.202	1.749	- 2.116	1.424	- 1.711	2.096	-1.320	1.615
1	-1.784	2.665	- 6.940	.149	8.258	3.292	- 4.970	3.898	- 4.724	2.589	- .188	4.769	-2.412	1.669
2	-1.719	2.206	- 1.232	-7.987	7.407	2.446	- 2.328	-.808	7.701	6.934	- 4.585	4.646	-3.655	-.151
3	-1.694	2.797	15.946	-3.752	-6.740	2.423	32.430	-6.986	12.212	9.008	- .397	6.583	-12.848	3.214
4	-1.671	-4.916	-9.423	3.522	-3.791	1.919	6.024	43.958	-4.404	10.491	- 6.308	6.751	12.616	4.345
5	-1.804	5.114	-6.205	4.355	-14.518	-1.428	-3.807	9.195	-13.609	9.993	10.228	- .277	.898	5.065
6	-2.131	3.901	-9.165	-3.707	4.816	6.886	-6.205	6.728	2.132	6.379	-2.664	- .942	3.364	5.227
7	-2.415	2.396	-4.614	-5.169	56.908	5.256	-2.873	34.958	-5.343	6.042	5.139	3.359	-5.805	5.299
8	-2.414	110.460	-9.863	4.453	-10.264	4.740	-40.876	23.832	-8.236	10.553	-8.313	10.112	-7.695	6.128
9	-2.357	9.120	-16.620	24.175	-15.26	11.335	-11.304	11.928	-6.320	4.608	-30.499	-12.344	-9.829	4.212
10	-2.239	6.722	-22.914	-62.662	-20.447	11.506	******	62.668	-5.472	58.460	-10.149	22.708	-10.696	5.455

S-array for the radio series

$f_m = (-1)^m \hat{\rho}(m)$.

TABLE 8

S11	S12	S13	S14
-4.734	.100	1.118	-2.119
1.004	-1.305	1.181	-2.482
1.415	-7.103	.398	-2.023
- .343	.306	1.539	-2.217
- .296	-17.485	1.217	-2.174
2.186	-1.718	1.433	- .988
-1.429	- .519	1.906	-1.586
2.412	-1.229	1.866	-59.006
- .988	.548	-.423	.420
.527	.145	-.407	-2.300
- .543	.296	-.464	- .500
.081	-.092	-.225	.659
.094	-3.005	-.290	.643
- .364	.348	-.197	.823
- .278	-3.313	-.661	.787
.764	- .069	-.413	.932
-3.204	.450	-.441	.422

S-array for the radio series after being transformed by 1-.9B.

$f_m = \hat{\rho}(m)$

TABLE 9

Estimated White Noise Variance .437271

Estimated AR Parameters 1.0066 -.1849 -.0346 .2085 -.2017 .1158 -.0697 .0821 -.0745 .0039 .2840 .1572 -.1267 -.2234

Root of Operator	Reciprocal of Root	Absolute Value of Root	Absolute Value of Reciprocal	Frequency	Period
(-.5424, 1.0460)	(-.3907, -.7534)	1.1783	.8487	.3261	3.0662
(-.5424, -1.0460)	(-.3907, .7534)	1.1783	.8487	.3261	3.0662
(.8693, .5105)	(.8553, -.5023)	1.0082	.9919	.0845	11.8327
(-1.0034, .8356)	(-.5885, -.4901)	1.3058	.7658	.3895	2.5675
(-1.0034, -.8356)	(-.5885, .4901)	1.3058	.7658	.3895	2.5675
(.5128, .8859)	(.4894, -.8455)	1.0236	.9769	.1665	6.0063
(.5128, .8859)	(.4894, .8455)	1.0236	.9769	.1665	6.0063
(-1.1873, .2043)	(-.8180, -.1407)	1.2048	.8300	.4729	2.1147
(-1.1873, -.2043)	(-.8180, .1407)	1.2048	.8300	.4729	2.1147
(.0120, 1.0466)	(.0110, -.9554)	1.0467	.9554	.2482	4.0295
(.0120, -1.0466)	(.0110, .9554)	1.0467	.9554	.2482	4.0295
(1.0553, .0577)	(.9448, -.0517)	1.0569	.9462	.0087	114.9376
(1.0553, - .0577)	(.9448, .0517)	1.0569	.9462	.0087	114.9376

AR(14) Yule-Walker fit to the Radio Series.

factors $1-(.94\pm.05i)B$ corresponding to the low frequency component of the data. Since the imaginary part is small, those two factors are close to being a double root of one. Noting Theorem 2 above, it is not surprising that, with only an estimate of $\rho(k)$ available, the constant behavior appears in the first column. That is often the situation when there is a strong low frequency oscillation in the data. It should be noted, however, that an oscillation is also apparent in the first column of the S-array which is contrary to the typical behavior of a process with a double root of one.

Examination of the factors given in Table 9 indicates that not all of the frequencies associated with the operator $1-B^{12}$ are present. It also indicates that some of those frequencies which are present are too far from the unit circle to be regarded as non-stationary. Thus the operator $1-B^{12}$ is clearly not part of the model for this data set. In fact use of that operator causes the low frequency component to be mistaken for something very near to a double root of one (that is, the model which results by first operating on the data by $1-B^{12}$ has one root of one and one real root slightly larger than one). That causes forecast functions for the model to have a nearly linear component to them, which is clearly an unreasonable forecast function for this data set.

As with the last example, we are now faced with the subjective decision as to whether a seasonal model is desired and, if so, which factors should be made seasonal.

The factors associated with the frequencies 1/12, 1/6, 1/4, and 1/123* seem the reasonable candidates to be made into seasonal factors. However, models including the factor associated with the

---

* Since the low frequency component is thought to be associated with sunspot activity we have given it the period estimated in Woodward and Gray [1978] for the sunspot series.

frequency 1/123 have forecast functions which are very unstable. To show the effect of choosing different seasonal models we consider the two models obtained taking the factors associated with 1/12 and with 1/12 and 1/6 as seasonal.

The S-arrays for the data filtered by $1-1.732B+B^2$ and by $(1-1.732B+B^2)(1-B+B^2)$ are given in Tables 10 and 11, respectively. The residual after operating by $1-1.732B+B^2$ appears well-modeled by an AR(12), as would be expected assuming the process to be well-modeled as an AR(14). The Yule-Walker fit is shown in Table 12 and is seen to be very similar to that shown in Table 9 (disregarding the factors already removed).

Examination of Table 11 indicates that the residual from the operator $(1-1.732B+B^2)(1-B+B^2)$ is well-modeled as an AR(13). The Yule-Walker fit to that model is given in Table 13. The higher order indicated here may be a consequence of Theorem 2 in that terms which were masked before removing the non-stationarities are now apparent.

Forecast functions of various lengths were calculated from a number of origins. The two models performed similarly in the cases considered. Two fairly representative forecast functions are given in Figures 4 and 5. We thus have two quite tenable models, both of which explain the data very well. Which model to use will depend on the uses to which the model will be put, and on what if any physical significance can be found in the extra parameters which were fit.

An important point of the above example is that the operator $1-B^S$ should not be used indiscriminately. The radio series is an example for which a cursory examination suggested the operator $1-B^{12}$ to be

TABLE 10

S9	S10	S11	S12	S13	S14
1.237	-1.285	.018	.715	-1.801	3.812
3.822	.620	- .755	.731	-1.306	.301
.897	.792	-5.444	.309	-1.657	1.113
1.496	.107	- .013	1.135	-1.624	-4.289
1.012	.031	-1.291	1.150	-1.901	-1.943
1.041	.894	- .954	1.056	- .895	1.690
.510	.011	- .559	1.180	-3.655	-3.696
.872	.541	- .554	1.443	7.162	7.757
-.120	.154	- .213	.250	- .241	.249
-.402	.005	- .154	.290	- .479	- .511
-.237	.151	- .153	.324	1.016	.875
-.327	.010	- .031	.183	- .412	.282
-.318	.030	.022	.163	- .558	.386
-.455	.158	- .161	.168	- .533	-1.160
-.336	.137	2.793	.434	- .568	- .126
-.984	-.498	.014	.250	- .421	.640
.019	.991	- .256	.254	- .338	- .270

S-array for the radio series after being transformed by $(1-1.732B+B^2)$. $f_m=\hat{\rho}(m)$.

TABLE 11

S10	S11	S12	S13	S14
.553	.309	2.400	-.855	.551
-.015	-3.177	3.430	-1.421	1.585
3.124	-3.191	-7.398	-1.235	4.456
.565	-1.079	.814	-1.595	-2.654
.596	-1.272	4.347	- .305	.269
.120	-1.756	-2.903	- .273	16.598
1.464	-1.626	- .906	-1.807	1.892
1.348	-3.320	.292	-1.646	-6.346
1.178	-1.163	2.255	-1.799	1.118
1.211	-56.977	-1.012	-2.111	4.165
3.583	3.490	3.434	-1.996	-11.104
.248	- .232	.249	- .284	.277
.470	- .501	- .205	- .327	.627
.491	11.729	.483	- .309	- .348
.151	- .234	- .078	- .212	.165
.139	- .814	.175	- .208	1.603
.081	- .488	.313	- .072	.081
.454	- .502	.841	- .080	- 4.589
.266	- .595	-1.060	- .615	.551
.284	- .345	.301	- .883	1.375
.007	- .431	6.170	- .994	4.511
.433	- .455	-1.406	- .467	- .296

S-array for the radio series after being transformed by $(1-1.732B+B^2)(1-B+B^2)$.

TABLE 12

Estimated White Noise Variance .443774

Estimated AR Parameters -.6222 -.3516 -.0406 .5644 .7075 .8096 .6345 .3147 -.1092 -.5030 -.4806 -.1732

Root of Operator	Reciprocal of Root	Absolute Value of Root	Absolute Value of Reciprocal	Frequency	Period
(-1.1574, .8057)	(-.5820, -.4051)	1.4103	.7091	.4032	2.4801
(-1.1574, -.8057)	(-.5820, .4051)	1.4103	.7091	.4032	2.4801
(.5101, .8891)	(.4855, -.8462)	1.0250	.9756	.1671	5.9843
(.5101, -.8891)	(.4855, .8462)	1.0250	.9756	.1671	5.9843
(- .5403, 1.0413)	(-.3926, -.7566)	1.1731	.8524	.3262	3.0659
(- .5403, - 1.0413)	(.3926, .7566)	1.1731	.8524	.3262	3.0659
(1.0556, .0634)	(.9439, -.0567)	1.0575	.9456	.0095	104.7397
(1.0556, -.0634)	(.9439, .0567)	1.0575	.9456	.0095	104.7397
(-1.2652, .1807)	(-.7746, -.1107)	1.2780	.7825	.4774	2.0946
(-1.2652, -.1807)	(-.7746, .1107)	1.2780	.7825	.4774	2.0946
(.0095, 1.0485)	(.0086, -.9537)	1.0485	.9537	.2486	4.0232
(.0095, -1.0485)	(.0086, .9537)	1.0485	.9537	.2486	4.0232

AR(12) Yule-Walker fit to the radio series after being transformed by $(1-1.732B+B^2)$.

TABLE 13

Estimated White Noise Variance .474858

Estimated AR Parameters -1.4782 -.7426 .6096 1.6976 1.5450 .5313 -.3647 -.4835 -.0887 .0620 -.1455 -.2820 -.1429

Root of Operator	Reciprocal of Root	Absolute Value of Root	Absolute Value of Reciprocal	Frequency	Period
(.9699, 1.0159)	(.4916, -.5150)	1.4046	.7119	.1287	7.7707
(.9699, -1.0159)	(.4916, .5150)	1.4046	.7119	.1287	7.7707
(-1.1307, .3733)	(- .7975, -.2633)	1.1908	.8398	.4492	2.2259
(-1.1307, - .3733)	(- .7975, .2633)	1.1908	.8398	.4492	2.2259
(- .4523, -1.0269)	(- .3592, .8156)	1.1221	.8912	.3160	3.1643
(- .4523, 1.0269)	(- .3592, -.8156)	1.1221	.8912	.3160	3.1643
(1.0536, .0650)	(.9456, -.0583	1.0556	.9473	.0098	101.9747
(1.0536, -.0650	(.9456, .0583)	1.0556	.9473	.0098	101.9747
(.0077, 1.0343)	(.0072, -.9668)	1.0343	.9668	.2488	4.0190
(.0077, -1.0343)	(- .0072, .9668)	1.0343	.9668	.2488	4.0190
(-1.1564, 0.0000)	(- .8648, 0.0000)	1.1564	.8648	.5000	2.0000
(- .8570, .8410)	(- .5944, -.5833)	1.2007	.8328	.3765	2.6561
(- .8570, - .8410)	(- .5944, .5833)	1.2007	.8328	.3765	2.6561

AR(13) Yule-Walker fit to the radio series after being transformed by $(1-1.732B+B^2)(1-B+B^2)$

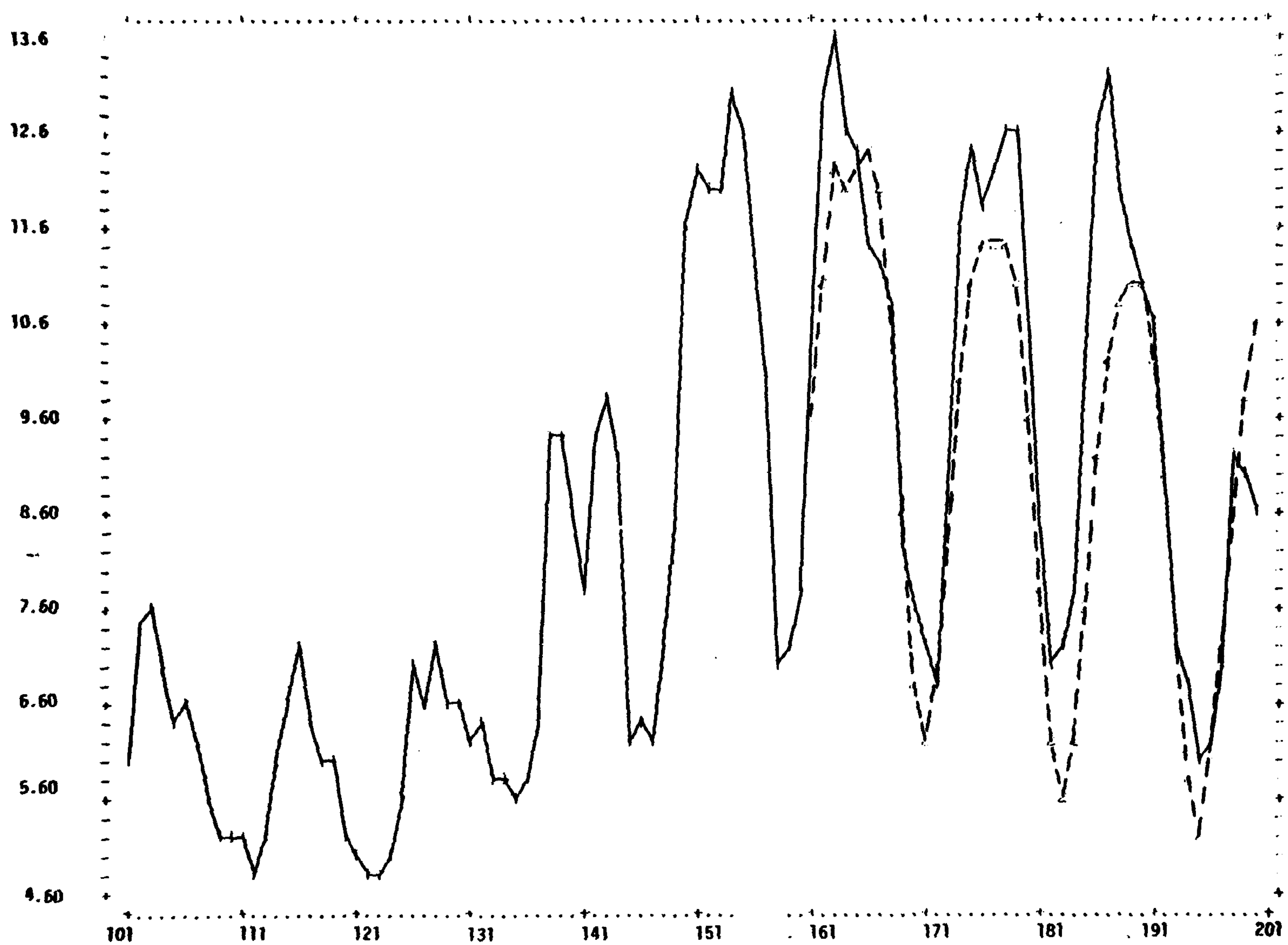

FIGURE 4. *Forecast function of* length 40 plotted against the realized values.

Model: $(1 - 1.732B + B^2)\ \phi_{12}(B)X_t = a_t$

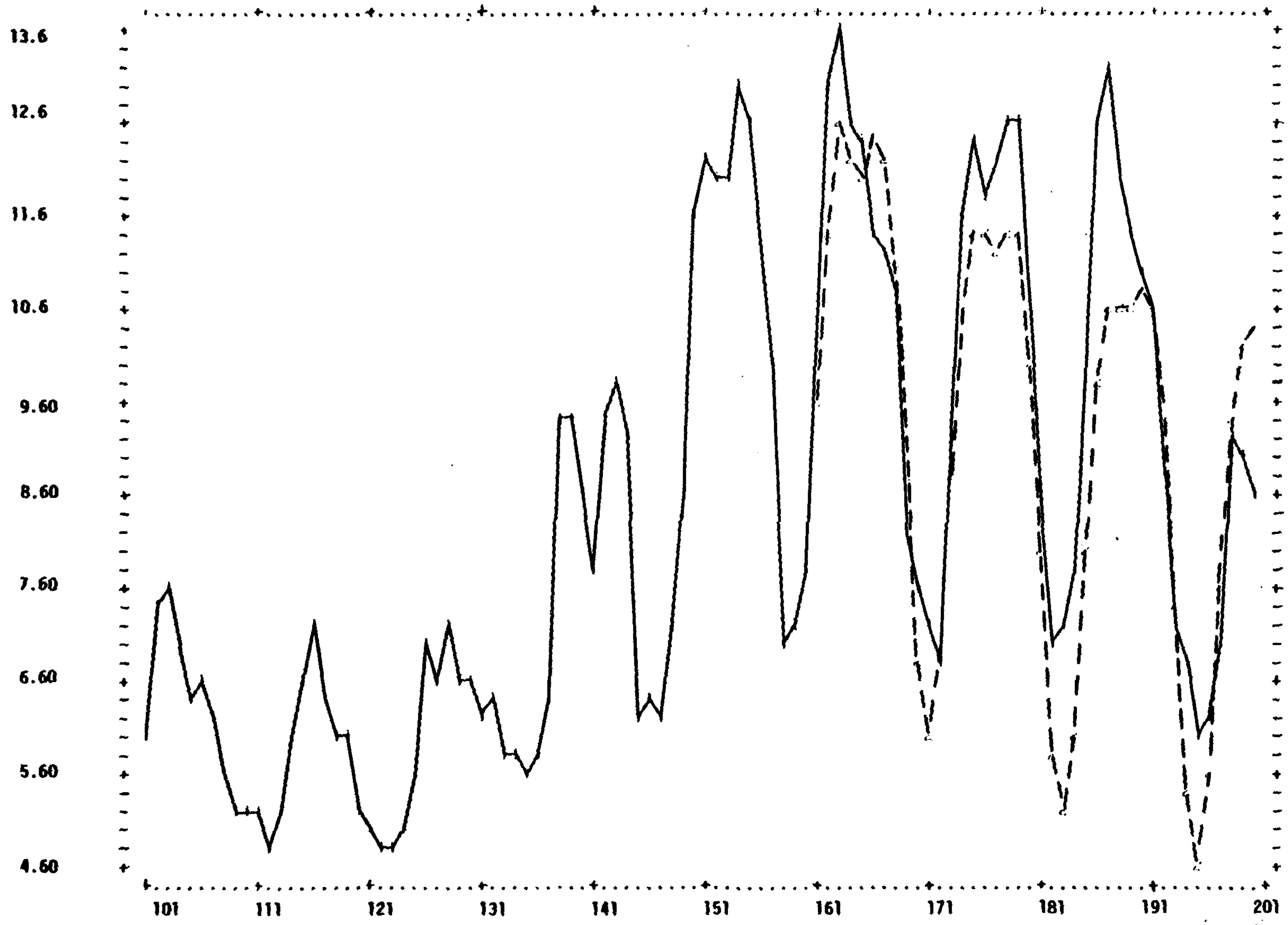

FIGURE 5. Forecast function of length 40 plotted against the realized values.

Model: $(1 - 1.732B + B^2)\ (1 - B + B^2)\ \phi_{13}(B)X_t = a_t$

appropriate. Further examination, however, showed that some of the frequencies associated with $1-B^{12}$ are not present in the data and that some of those present are clearly not on the unit circle. In fact as already pointed out the use of the operator $1-B^{12}$ on the radio series causes one of the most salient points of the data set--the low frequency oscillation--to be mistaken for essentially a double root of one. So if, for instance, the purpose of the analysis was to investigate the fitted model for evidence that sunspot activity influences radio transmission that evidence has been badly obscured if not lost.

ACKNOWLEDGEMENT: I wish to thank Professor Henry Gray for the numerous helpful comments he has made on earlier drafts of this paper and for the many discussions we have had on this and related topics.

REFERENCES

Box, G.E.P. and Jenkins, G. M. (1976), Time Series Analysis: Forecasting and Control.

Findley, David, (1980), Personal Communication.

Gray, H. L., Kelley, G. D., McIntire, D. (1978), "A New Approach to ARMA Modeling", Communications in Statistics, B7, pp. 1-77.

Gray, H. L., and Woodward, W. A. (1980), "Application of S-arrays to Seasonal Data", Proceedings of the 1979 Tulsa Time Series Symposium.

Hart, J. and Gray, H. L. (1980), "Modeling Seasonal ARMA Processes," ONR Tech. Report No. 140, SMU Department of Statistics.

Parzen, E. and Pagano, M. (1970), "An Approach to Modeling Seasonally Stationary Time Series," Journal of Econometrics.

Quinn, Berry (1980), "Limiting Behavior of Autocorrelation Function of ARMA Processes as Several Roots of the Characteristics Equation Approach to the Unit Circle," Communications in Statistics, B9, pp. 195-198.

Siddiqui, M. M. (1962), "Some Statistical Theory for the Analysis of Radio Propagation Data," Journal of Research NBS, 66D (Radio Propagation), pp. 571-580.

Woodward, W. A. and Gray, H. L. (1979), "On the Relationship Between the R and S-Arrays and the Box-Jenkins Method of ARMA Model Identification," ONR Tech. Report No. 134, SMU Department of Statistics.

Woodward, W. A. and Gray, H. L. (1978), "New ARMA Models for Wolfer's Sunspot Data, "Communications in Statistics, B7, pp. 97-115.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 141	2. GOVT ACCESSION NO. AD-A097	3. RECIPIENT'S CATALOG NUMBER 516
4. TITLE (and Subtitle) The Determination of Seasonal ARMA Models		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER 141
7. AUTHOR(s) Michael Morton		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0439
9. PERFORMING ORGANIZATION NAME AND ADDRESS Southern Methodist University Dallas, Texas 75275		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042 280
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, Va. 22217		12. REPORT DATE October 1980
		13. NUMBER OF PAGES 32
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any purposes of the United States Government.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Gray, Kelley, and McIntire (1978) have described a method for determining the order of an ARMA process and for identifying roots of the characteristic equation on or near the unit circle. In this paper, we will demonstrate how that approach can be utilized in modeling seasonal data.